

Bankruptcy Prediction with Support Vector Machines: An Application for German Companies

Master Thesis Submitted to

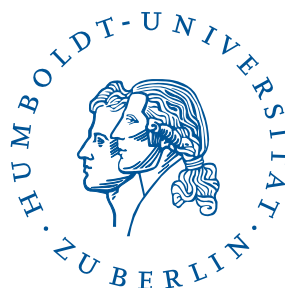
Prof. Dr. Wolfgang K. Härdle

Linda Hoffmann

Ladislaus von Bortkiewicz Chair of Statistics

C.A.S.E.- Centre for Applied Statistics and Economics

Humboldt-Universität zu Berlin



by

Ceren Önder

(521872)

in partial fulfillment of the requirements

for the degree of

Master of Economics and Management Science

Berlin, December 24th, 2010

Acknowledgment

I would like to express my deep and sincere gratitude to my supervisor, Prof. Dr. Wolfgang K. Härdle, for his important support and encouragement throughout this work. Many thanks also to my advisor, Linda Hoffmann, for her constructive criticism and insightful comments in all the time of writing my thesis which definitely enabled me to develop a deep understanding of the subject.

I would like to extend my thanks to my brother, Asim Önder, who has made available his kind support in a number of ways. I am indeed grateful to him for spending his time in helping me to finish this work with the best possible results. My sincere thanks also go to my best friend, Damla Kardes, for being with me all the time from the initial to the final level of this work. Her loving support has been of great value.

Last but not the least, I would like to thank my parents, Gonca and Süleyman Önder, for supporting me spiritually throughout my life. Without their understanding and encouragement it would have been impossible for me to accomplish this work.

Ceren Önder

Abstract

This work investigates the application of support vector machines (SVMs) for prediction of German companies' failure that is based on 24 financial ratios being grouped into four categories, namely profitability, leverage, liquidity and activity. SVMs, one of the efficient supervised learning methods, can be applied to classification or regression. In our framework, we will use SVMs as a classifier to discriminate solvent and insolvent companies. To select the appropriate model composed of the most powerful financial ratios in classification both forward and backward stepwise selection techniques are utilized and the derived results are compared through their accuracy ratios. Furthermore, with the use of the same financial ratio data we predict companies' probability of default via logistic regression. The selection of the regression model is based on Akaike's information criterion. Then, the best established results of the SVM and logistic models are assessed via their Lorenz curves to figure out which model is more successful method to extract useful information from financial ratios for default prediction. The results release that the SVM model outperforms logistic regression.

Keywords: Financial Ratio, Default Probability Prediction, Support Vector Machines, Logistic Regression

JEL classification: G33, C45 ,C14, C44

Zusammenfassung

Diese Arbeit untersucht die Anwendung von Support Vektor Machines (SVMs) zur Vorhersage der Insolvenz von deutschen Unternehmen. Die Vorhersage basiert auf 24 finanziellen Kennzahlen, die in vier Kategorien unterteilt sind: Profitabilität, Fremdfinanzierung, Liquidität und Aktivität. SVMs, eine der effizientesten geordneten Lernmethoden, können zur Klassifikation oder zur Regression dienen. In unserem Fall, werden wir SVMs als Klassifikator benutzen, um solvente und insolvente Unternehmen zu diskriminieren. Um das geeignetste Model, das aus den aussagekräftigsten finanziellen Kennzahlen zusammengesetzt ist, auszuwählen, werden sowohl vorwärts, als auch rückwärts schrittweise Selektionstechniken verwendet, und die hergeleiteten Resultate werden durch ihre Genauigkeitsquote. Des Weiteren nutzen wir die gleichen finanziellen Kennzahlen, um die Insolvenzwahrscheinlichkeit der Unternehmen mit einer logistischen Regression vorherzusagen. Die Wahl des Regressionsmodels basiert auf Akaikes Informationskriterium. Die jeweils besten Resultate der SVM und des logistischen Models werden anhand der Lorenzkurve verglichen, um herauszufinden, welches Model erfolgreicher darin ist, wesentliche Informationen für Insolvenzhervorsagen aus den finanziellen Kennzahlen zu ziehen. Die Resultate des SVM-Models übertreffen diejenigen der logistischen Regression.

Schlagwörter: Finanziellen Kennzahlen, Insolvenzhervorsagen, Support Vektor Machines, Logistische Regression

JEL Klassifikation: G33, C45, C14, C44

Contents

1	Introduction	1
2	The effect of Basel Accords on the financial system	5
3	Bankruptcy prediction with Support Vector Machines (SVM)	9
3.1	SVM: Algorithm	10
4	Bankruptcy prediction with logistic regression	21
5	Data description	23
5.1	Data cleaning and model selection	23
6	Empirical Results	29
6.1	Results on the SVM model	29
6.2	Results on logistic regression	33
6.3	Comparison of the SVM and logistic models	38
7	Conclusion	41

1 Introduction

It is required for any company, private or public, to have funds to run. In the case of insufficient means to meet liabilities, the management of the company will have to liquidate the company's assets, but when there are not enough funds to repay all debtors fully, the declaration of bankruptcy can happen. Since corporate liabilities have default risk, for the banking industry and financial institutions it is essential to discriminate the non-defaulting companies from defaulting ones. Through a good discrimination they can make lending decisions and pursue better pricing strategies and so cover against the counter party risk. To do that these institutions develop their own models relying substantially on statistical tools. Nonetheless, it is not easy to model the relevant risk accurately so that even small miscalculations of the risk can impede the profitability of lending. Having an accurate model to predict corporate failure is of great interest to a variety of parties, such as, insurers, factors, financial analysts and academics. Because of the fact that unexpected realizations of default risk can destabilize and destroy lending institutions, it will always be crucial to improve bankruptcy prediction accuracy.

The problem of prediction of company failure is not a new topic. To the contrary, it has already been analyzed since the 19th century. Similarly, the question of "How can financial data contribute to our understanding of why some firms go into bankruptcy?" has been more than a century old. Ramser and Foster (1931), Fitzpatrick (1932) can be counted as the first scientists applying financial ratios so as to predict company bankruptcy. Nonetheless, Aziz et al. (1988) criticized the models which built bankruptcy prediction on financial ratios. They claimed that corporate bankruptcy is highly related to firm valuation and that is why a cash flow model can predict corporate failure more accurately. Similarly, Sierra and Bechetti (2003) investigated non-financial explanatory variables in order to predict company failure, such as the degree of relative firm inefficiency, customers' concentration and the strength and proximity of competitors.

In the late 1960s, discriminant analysis (DA) was introduced as the first applied statistical technique to bankruptcy prediction. The publications of Beaver (1966) and Altman (1968) are the first works using this method. Afterwards, a variety of models have been developed via using techniques such as logit, probit, hazard models, recursive partitioning and neural networks. Beaver (1966) introduced the univariate DA and provided empirical verification that certain financial ratios, most importantly cash flow to total debt ratio, gave statistically significant signals well before actual business failure occurs. Furthermore, in predicting company failure he stressed that accounting data are crucial and he found that market-value variable and financial ratio are equally reliable. Altman (1968) improved Beaver's univariate analysis through developing a discriminant function which allows for simultaneous consideration of several ratios in a multivariate analysis, namely multivariate DA, in order to predict company default and he concluded that some of the ratios he used outperformed Beaver's cash flow to total debt ratio. Altman asserted that an analysis with just one financial ratio at a time is not sufficient in failure prediction. For example, according to him a firm posing a poor picture in terms of profitability

may be regarded as having high risk of bankruptcy. Nevertheless, other financial indicators such as above average liquidity can change this adverse situation. Later on, Altman et al. (1977) developed Z-score model for predicting corporate bankruptcy which became a control measure for the financial distress status of the companies. The Z-score is calculated by a linear combination of several common business ratios which are weighted by coefficients. Thanks to its simplicity, it is still widely used in bankruptcy prediction.

Beaver and Altman used financial ratios in order to predict failure of medium and large asset size firms while ignoring small businesses due to the difficulties in acquiring the data. In contrast, Edmister (1972) claimed that bankruptcy is a more common phenomenon for small companies and analyzed the prediction of small business failure by using financial ratios.

Linear probability and multivariate conditional probability models, Logit and Probit, were utilized in the corporate failure prediction literature of the late 1970s. The odds of a corporate failure was estimated with the probability via these methods. Martin (1977) introduced "Early warning of bank failure", wherein a logit regression approach is implemented. Later, Moro et al. (1980)'s work of "Financial ratios and the probabilistic prediction of bankruptcy" based on logit model was published, where he used a large sample, 105 bankrupt firms and 2,058 non-bankrupt firms, being different from the previous studies. The drawback of this study is that he disregarded the statements after bankruptcy. Some of the other researchers applying logistic and probit model to bankruptcy analysis can be counted as Wiginton (1980), Zavgren (1983) and Zmijewski (1984). As to the other statistical methods used to predict company failures, there are namely gambler's ruin model (Wilcox (1971)), option pricing model (Merton (1974)), recursive partitioning (Frydman et al. (1985)), neural networks (Tam and Kiang (1992)) and rough sets (Dimitras et al. (1999)).

Wilson et al. (1999) asserted that the contribution of the ability and willingness of a firm to pay its creditors to credit analysis is highly important. In their model, they used a linear combination of financial ratios grouped into categories of liquidity, leverage, business activity and profitability along with variables obtained from non-financial and payment behavior information, such as company age, company size and industry type. They concluded that the inclusion of a history of payment behavior and non-financial data to the models can improve the bankruptcy prediction.

In 2001, Shumway criticized single-period models so-called static models and claimed that through applying these models one can only produce probabilities that are biased and inconsistent. According to him, firms change through time and static models do not control firms' risk for each time period. Hence, Shumway and Hope (2001) suggested a hazard model, a multi-period assessment, including time-varying explanatory variables so as to catch the changing company status over time. Furthermore, Glennon and Nigro (2005) suggested a hazard or survival analysis. Similarly, Duffie et al. (2007) predicted corporate distress via multi period model, and they also incorporated the dynamic of firm-specific and macroeconomic covariates. They found the strong dependence of company health status on the current state of the economy and the current leverage of a company.

As to the SVM model introduced by Cortes (1995), it is a supervised learning method based on mathematical programming theory which can be applied to both classification and regression analysis to model and predict corporate failure. Fan and M. (2000), Min and Lee (2005) and Shin et al. (2005) can be given as examples for the previous studies using this method in bankruptcy prediction. In this work, through using Gaussian radial basis function (RBF) as a kernel trick we will use SVMs as a nonlinear classifier to separate solvent and insolvent com-

panies. Thanks to kernel trick we will transform the financial ratios into a higher dimensional space where linear separation of the companies is possible.

The remainder of this paper is structured as follows. In Section 2, the important effect of Basel Accords on the financial system is described. The use of the SVM model in bankruptcy prediction is then explained in detail with its related algorithm in Section 3. In Section 4, we present the logistic regression in corporate failure prediction. We render how we clean data and then select the appropriate model for the SVM in Section 5. We provide the empirical results in Section 6 and we conclude in Section 7.

2 The effect of Basel Accords on the financial system

It is also required to glance at the importance of bankruptcy analysis in a broader sense. For a society, it is crucial to have an efficient bankruptcy system. When the efficiency of an applicable bankruptcy system surges, interest rates decline, which leads to increase in creditor payoffs and decrease in cost of debt capital. This fall in cost of debt capital raises the firms' share of good state returns. With an efficient bankruptcy system the firms have more investment incentives, since the firms decide to carry out projects in an effort to maximize net expected profits which soar when the interest rate decreases. All in all, an efficient bankruptcy system has a major role in stabilizing financial markets which also gives rise to stable inflation and sustainable growth. It is required to have clear and consistent rules which govern the process aiming to have a more stable financial system. In this respect, the role of the Basel Committee on Banking Supervision (BCBS) being established as an international committee by the Bank for International Settlements is of great importance for financial system, because it aims to strengthen supervisory and risk management practices globally.

BCBS disclosed Basel I, the first Basel Capital Accord, in 1988 which was enforced by law in G-10 countries in 1992. Basel I mainly focused on credit risk. The ultimate aim was to determine capital standards and to ensure that banks have a sound amount of capital. It required the banks with international presence to possess a minimum capital to risk weighted asset ratio of 8%. Holding a certain amount of capital is crucial because of the fact that capital of financial institutions is the last guard against bank insolvency.

Because of the developments in financial innovations and risk management BSBC started in 2000 working on Basel II to create more comprehensive and risk sensitive set of guidelines to capital adequacy. Within this Accord being revised in 2004 one-size-fits-all methodology for computing capital ratio was transformed to a set of three choices; the standardized approach, the internal ratings-based approach (IRB) and the advanced internal ratings-based approach (AIRB). The difference from one-size-fits-all methodology is that IRBs allow the participating financial institutions to utilize their own internal estimates of risk components in determining the capital requirement to guard against financial and operational risk. Importantly, BSBC added operational risk into the capital to assets ratio. In short, the aim of Basel II is to set an international standard to determine the amount of capital that banks require to hold in order to protect themselves against financial and operational risks. Such an international standard can guard international system from major banks collapse. When a bank puts aside capital reserves in parallel to the risk in its lending and investment activities, this will lead to economic stability. The greater this risk, the greater the amount of capital the bank requires to hold in order to protect its solvency. It is obvious that, this point differs from Basel I which required banks to have minimum capital amount of 8%. The difference between two Basel Accords can be seen in Table 2.2.

Rating Class (Moody's)	Four year PD (%)
Aaa	0.02
Aa2	0.05
Aa3	0.11
A1	0.21
A2	0.38
A3	0.59
Baa1	0.91
Baa2	1.32
Baa3	2.62
Ba1	4.62
Ba2	7.48
Ba3	10.77
B1	15.24
B2	19.94
B3	26.44
Caa1	35.73
Caa2	72.87
Caa3	48.27
Ca	100.00
C	100.00

Table 2.1: PDs used in rating structure finance securities. Source: Moody's Rating Methodology (2006)

As a new update to the Basel Accords, in September 2010 BCBS announced the Basel III which was confirmed by the G-20 in November 2010 at Seoul G-20 leaders' summit. New Basel Accords aim to strengthen capital and liquidity requirements in response to the last global financial crisis stemmed from the mortgage and credit crunch in the US. In addition, the required capital conversation buffer will also be increased which is crucial to ensure that banks have capital buffer that can cover losses during the periods of financial and economic distress. Introduction of a global liquidity standard and new capital buffers with tightened minimum common equity requirement are important financial sector reforms supporting the economic growth. Obviously, the goal is to make the financial firms stronger against future financial crises.

A company's bond rating reflects its financial strength which is negatively correlated to the cost of debt. Cost of debt is the yield to maturity on the company bonds or simply the cost of borrowing money. Therefore, the higher the company rating, the higher company's ability is to make repayments of interest and principal; in other words, the cost of debt will be lower for a high rated company. The most important credit rating agencies are Moody's, Standard and Poor's and Fitch IBCA that publish ratings in the US. Moody's rating methodology disclosed in 2006 for four year PDs is given for an example in Table 2.1.

As the Basel Accords emphasize, it is important for a bank or any other financial institution to quantify the risk of insolvency of the borrower. At this point, which rating method is the most accurate one to discriminate the low-risk and high-risk companies becomes important. If financial institutions can find the adequate model, they can invest in low-risk companies and

Rating Class (S&P)	One-year PD (%)	Capital Requirements (%) (Basel I)	Capital Requirements (%) (Basel II)
AAA	0.01	8.00	0.63
AA	0.02 – 0.04	8.00	0.93 – 1.40
A+	0.05	8.00	1.60
A	0.08	8.00	2.12
A-	0.11	8.00	2.55
BBB	0.15 – 0.40	8.00	3.05 – 5.17
BB	0.65 – 1.95	8.00	6.50 – 9.97
B+	3.20	8.00	11.90
B	7.00	8.00	16.70
B-	13.00	8.00	22.89
CCC	> 13	8.00	> 22.89
CC		8.00	
C		8.00	
D		8.00	

Table 2.2: Rating grades and capital requirements. Source: Damodaran (2002) and Fuser (2002). The authors estimated the figures in the last column for a loan to a small and medium enterprise with a turnover of 5 million Euros due 2.5 years through using the data from column 2 and the recommendations of the Basel Committee on Banking Supervision, 2003.

then have the chance to decrease capital requirements.

3 Bankruptcy prediction with Support Vector Machines (SVM)

As mentioned in the previous part, the calculation of the likelihood that a company may go bankrupt is highly important for the creditors. On this account, finding the best separation model depending on company bankruptcy predictions is of great consideration. As of the late 1980s, machine learning techniques, one of the artificial intelligence techniques, have been applied to bankruptcy prediction. The SVM being a decision-based prediction algorithm is a relatively new one among these techniques. SVMs can be thought as the most promising system when compared to other statistical techniques, since it has many affirmative features and advanced generalization performance on the results. As to the comparison of the SVM with the other machine learning techniques, its advantages stand out clearly. For instance, Artificial Neural Network (ANN) is the most frequently used learning technique in literature due to its simplicity and high performance in prediction, but this technique has important drawbacks. ANNs require a large amount of training data so as to estimate the distribution of input pattern. Besides, due to their over-fitting nature they have serious difficulties in generalizing the results. On the other hand, the SVM can overcome this over-fitting risk through pulling away the distance between solvent and insolvent companies. Moreover, ANNs depend more on heuristics in comparison to SVMs.

In the application of the SVM to corporate failure prediction, the goal is to find the optimal hyperplane that separates the analyzed companies into two groups of solvent and insolvent. If it is possible to perform the separation of the cluster of points by a straight line, it will be very easy. Unfortunately, in general this is not the case. At this point the superiority of the SVM arises because through applying the SVM, classification becomes possible in the case of nonlinear separable data. But how? Through transforming the data points into a higher dimensional space with the help of a kernel function (Hastie and Friedman (2001)). Hence, we can conclude that the application of the SVM yields a linear model constructed in a higher dimensional input space which represents the nonlinear decision boundary in the original space. This feature is a very powerful advantage of the SVM in bankruptcy prediction in comparison with the other statistical methods, such as classical DA, logit and probit requiring linearly separable data.

In the new higher dimensional space, the mapped input vectors are separated into two categories of data by an optimal plane being called "feature". Here, the question is what a good separation is. It can be derived by the hyperplane that has the largest distance to the nearest training data points of any class. This distance is called "margin". The larger the margin the better the separation between decision classes is. A set of features that describes one case is called vector and the vectors that are closest to the maximum margin hyperplane are called support vectors.

All in all, the aim of the SVM in bankruptcy prediction is to maximize the margin between support vectors in order to get the best separation between bankrupt and non-bankrupt companies.

3.1 SVM: Algorithm

As a supervised learning method SVMs allow for analyzing the training set, a set of labeled observations, in order to predict the labels of unlabeled future data given by the test set. Hence, the aim is to achieve some function that yields the relationship between observations and their labels that makes it possible to arrange new observations into existing classes. In this section, training algorithms for SVMs will be analyzed.

Generalization error in supervised learning, which gives the degree of misclassification, determines how good the algorithm will be in classifying the future data. In this sense, this concept can be thought as loss. Suppose the data points are $x \in X$ being financial ratios, true labels are $y_i \in \{-1, +1\}$ representing the bankruptcy ($y_i = 1$) and non-bankruptcy ($y_i = -1$) situations of the companies, and $f(x)$ is the classifying prediction function which derived from the available set of measurable functions F . Hence, misclassification will be given by the situations when $f(x) \neq y$. In order to understand how well the SVM algorithm performs on the classification, expected risk minimization principle (ERM), the expected value of loss under the true probability measure, is utilized given by:

$$R(f) = \int \frac{1}{2} |f(x) - y| dP(x, y), \quad (3.1)$$

There is an unknown dependence between x and y , and there is no information about the joint probability functions $P(x, y)$ in practical applications (Wang (2005)). To estimate this relevant risk the only information about the distribution can be obtained from the training set $\{(x_i, y_i)\}_{i=1}^n$, a set of data taken out from the entire dataset to discover the predictive relationship between labels and financial ratios, but this solution gives rise to an ill-posed problem Tikhonov and Arsenin (1997). How can we know how good the prediction is without knowing the true distribution?

According to an assumption of Cortes (1995) there is an access to a sequence of independent random variables all of which are derived from the true distribution. Importantly, the observations are assumed to be independent and identically distributed. Related to these assumptions there are two ways for approximating to the unknown true distribution; "minimizing the empirical risk" and "calculating the Vapnik-Chervonenkis (VC) bound".

The empirical risk minimization that can be used as a surrogate for expected risk minimization can be calculated from:

$$\hat{R}(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} |f(x_i) - y_i|, \quad (3.2)$$

As it is seen, the average value of loss over the training set simply gives the empirical risk.

It is crucial to stress that if F incorporates too many candidate functions or the training set is not large enough, empirical risk minimization paves the way for unsuccessful generalization and high variance. In this case, learning algorithms can just memorize the training examples with a poor generalization which is called over-fitting. If F is not too large, the increase in n will lead to an approximation of both risk minimizations to each other. It is obvious that the results of the expected and empirical risk minimization are not required to be equal as can be seen in Figure 3.1.

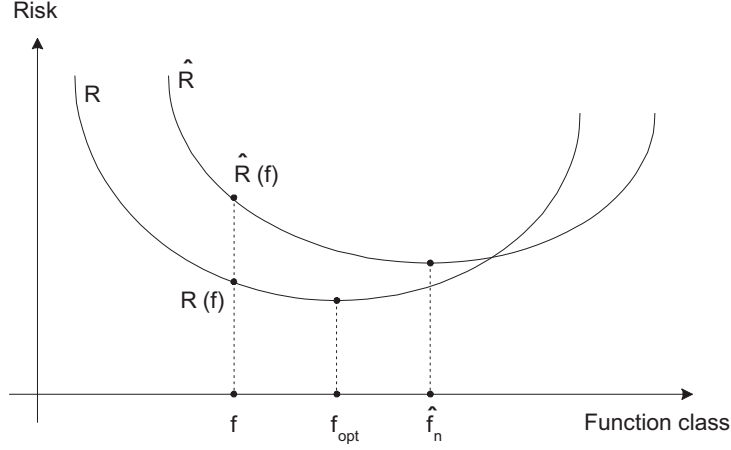


Figure 3.1: The minima f_{opt} and \hat{f}_n of the expected (R) and empirical (\hat{R}) risk functions generally do not coincide. Source: Moro, Hoffmann and Härdle (2010)

$$f_{opt} = \arg \min_{f \in F} R(f), \quad (3.3)$$

$$\hat{f}_n = \arg \min_{f \in F} \hat{R}(f), \quad (3.4)$$

As put before, the other way in statistical learning theory to approximate the distribution of $P(x, y)$ is calculating the VC bound introduced by Vapnik and Chervonenkis (1971) which represents distribution independent bounds on the generalization performance of a learning machine. Through VC dimension we predict a probabilistic upper bound for the expected risk on the test error of a classification model which holds with a certain probability $1 - \eta$. The VC bound can be written as (Vapnik and Chervonenkis (1971)):

$$R(f) \leq \hat{R}(f) + \phi\left(\frac{h}{n}, \frac{\ln(\eta)}{n}\right). \quad (3.5)$$

For a linear indicator function $g(x) = \text{sign}(x^\top w + b)$:

$$\phi\left(\frac{h}{n}, \frac{\ln(\eta)}{n}\right) = \sqrt{\frac{h \left(\ln \frac{2n}{h} \right) - \ln \frac{\eta}{4}}{n}}, \quad (3.6)$$

where h is the VC dimension of the classification model which determines the complexity of a classifier function and n is the size of the training set. h can be defined as:

$$h \leq \min\{(r^2 \|w\|) + 1, n + 1\} \quad (3.7)$$

where r is the radius of the smallest sphere composing of data.

As to SVMs, the VC dimension of the function set F in a d -dimensional space is h if some

3 Bankruptcy prediction with Support Vector Machines (SVM)

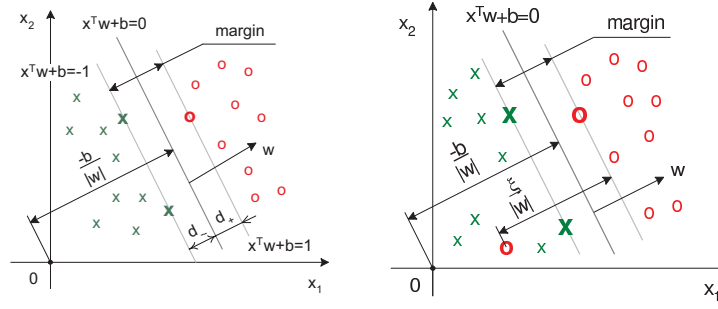


Figure 3.2: The margin in a linearly separable (left) and non-separable (right) case. Crosses and circles are solvent and insolvent companies, respectively. Source: Moro, Hoffmann and Härdle (2010)

function $f \in F$ can shatter h objects $\{x_i \in \mathbb{R}^d, i = 1, \dots, h\}$ in all 2^h possible layouts and no set $\{x_j \in \mathbb{R}^d, j = 1, \dots, q\}$ exists where $q > h$ that satisfies this property. For instance, suppose that a straight line is the classification model which has to divide data into two groups and suppose that there are sets of 3 points on a plane ($d = 2$) that can be shattered via this model in $2^h = 2^3 = 8$ ways. Nevertheless, in the case of set of 4 points it is not possible to separate one of these two subsets from the other. Thus, the VC dimension of this classifier in a two-dimensional space is three.

Additionally, the functional for the VC bound seen in (3.5) should be further analyzed. Here, VC dimension h is a parameter to control the complexity of the classifier function determining how complicated the classification model can be. Importantly, the second term $\phi\left(\frac{h}{n}, \frac{\ln(\eta)}{n}\right)$ represents regularization term introducing a penalty for the excessive complexity of a classifier function. In short, the expected VC-dimension will yield the complexity. Lower h will give rise to a larger margin which leads to a good generalization for the future data, but the misclassification can be higher. Hence, it can be said that there is a trade-off between the misclassification on the training set and the complexity of the classifier function. The SVM aims to construct a model having a minimized VC dimension. (Wang (2005))

The next step in analyzing the SVM is maximizing the margin, the distance of the hyperplane to the closest point in the dataset. This distance shows how well the data is separated. As expected, the wider the margin the clearer is the separation because the ultimate goal is to generalize the separation for the future data as well. Put another way, when the hyperplane is as far away as possible from both of the classes, the generalization is expected to become better. To find the maximum margin we need to solve an optimization problem.

There are two different ways to treat the SVM problem. The first one is the hard margin SVM introduced by Vapnik and Lerner (1963), where the dataset is assumed to be linearly separable by the maximum margin which leads to perfect classification of every data point that can be visualized on the left-hand side in Figure 3.2. However, as mentioned before this is not always the case as seen on the right-hand side of the same figure. Soft margin is developed by Mangasarian and Bennett (1992) for data being not linearly separable which allows some degree of misclassification of data points.

In the linearly separable case the classification hyperplane of the SVM is:

$$x^\top w + b = 0, \quad (3.8)$$

where w is the weight or slope with a dimension of $d \times 1$. Here, d will give the number of characteristics of the companies used so as to classify them into solvent and insolvent groups. In this work financial ratios will be used for this purpose. Hence, we have d financial ratios. x_i is a vector of dimension $d \times 1$ representing the financial ratios of company i . Finally, b is a scalar that gives the location parameter, the so called threshold.

As put before, when data are linearly separable, the points can be accurately classified with a maximum margin meaning that all observations should satisfy the following constraints:

$$\begin{aligned} x_i^\top w + b &\geq 1 \text{ for } y_i = 1, \\ x_i^\top w + b &\leq -1 \text{ for } y_i = -1, \end{aligned}$$

It is possible to write these two constraints together:

$$y_i(x_i^\top w + b) - 1 \geq 0 \quad i = 1, 2, \dots, n \quad (3.9)$$

It is obvious that we have two groups that can be shown as $+1(-1)$. We turn back to the left side of Figure 3.2 again, where $x_i^\top w + b = \pm 1$ are the canonical hyperplanes being parallel to each other. Then $d + (d-) = 1/\|w\|$ gives the shortest distance between the observations being in the class in question $+1(-1)$ and the separation hyperplane. So the width of the margin is given by $2/\|w\|$. Thus, maximizing the margin means minimizing the Euclidean norm $\|w\|$ or its square $\|w\|^2$.

This margin maximization optimization problem can be expressed by a primal Lagrangian as:

$$\min_{w, b, \xi_i} \max_{\alpha_i} L_P = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i \{y_i (x_i^\top w + b) - 1\}, \quad (3.10)$$

where α_i is a Lagrange multiplier.

The Karush-Kuhn-Tucker (KKT) first order optimality conditions are:

$$\begin{aligned} \frac{\partial L_P}{\partial w_k} = 0 &\iff w_k - \sum_{i=1}^n \alpha_i y_i x_{ik} = 0, \quad k = 1, \dots, d \\ \frac{\partial L_P}{\partial b} = 0 &\iff \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L_P}{\partial \alpha_i} = 0 &\iff y_i (x_i^\top w + b) - 1 \geq 0 \quad i = 1, \dots, n \\ &\alpha_i \geq 0 \\ &\alpha_i \{y_i (x_i^\top w + b) - 1\} = 0 \end{aligned}$$

3 Bankruptcy prediction with Support Vector Machines (SVM)

where $\alpha_i \geq 0$ is the sign condition on the multiplier and $\alpha_i \{y_i (x_i^\top w + b) - 1\} = 0$ represents the complementary slackness.

Then we need to introduce these KKT conditions to the primal problem. We rewrite (3.10) as:

$$\frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i x_i^\top w - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i,$$

We can express $\sum_{i=1}^n \alpha_i y_i x_i^\top w$ with the elements of the matrices as:

$$\sum_{k=1}^d \sum_{i=1}^n \alpha_i y_i x_{ik} w_k \quad (3.11)$$

As a reminder, d gives the number of financial ratios in this work. Next, we insert $w_k = \sum_{j=1}^n \alpha_j y_j x_{jk}$ to (3.11) and get:

$$\sum_{k=1}^d \sum_{i=1}^n \alpha_i y_i x_{ik} \left(\sum_{j=1}^n \alpha_j y_j x_{jk} \right) = \sum_{k=1}^d \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i x_{ik} \alpha_j y_j x_{jk}$$

Now we can write $\sum_{k=1}^d \sum_{i=1}^n \sum_{j=1}^n x_{ik} x_{jk}$ part in matrix terms with $x_i^\top x_j$, and what we have becomes:

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j$$

Importantly, we cancel $b \sum_{i=1}^n \alpha_i y_i$ through substituting one of the KKT conditions (Gale and Tucker (1951)) saying $\sum_{i=1}^n \alpha_i y_i = 0$. Finally, we write (3.10) again.

$$\frac{1}{2} \|w\|^2 - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j + \sum_{i=1}^n \alpha_i, \quad (3.12)$$

We know that the Euclidean norm is $\|w\| = \sqrt{\sum_{k=1}^d w_k^2}$ and so we can say that the first term of (3.12) gives the half of the sum of squared weights. Hence, together with the second term which is the sum of the squared weights, the primal problem can be written as the dual problem:

$$\max_{\alpha_i} L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j \quad (3.13)$$

$$\text{s.t. } \alpha_i \geq 0, \quad (3.14)$$

$$\sum_{i=1}^n \alpha_i y_i = 0. \quad (3.15)$$

It is crucial to note that this optimization problem is convex that both primal and dual problems will yield the same result.

However, as stated before in general this perfectly linearly separation cannot be the case in real life applications where very high dimensional problems are possible. Therefore, it is required now to switch to more general case - linearly non-separable case.

Due to some reasons such as outliers in the dataset it is highly possible to have some data points that fall inside of the margin zone or are on the incorrect side of the margin according to classification. In this situation, it is necessary to have some cost for each misclassification which is determined by how far the data point is from maintaining the margin requirement. Because of this, the inequalities utilized for classification can be rewritten as below which holds for all n data points of the training set:

$$x_i^\top w + b \geq 1 - \xi_i \quad \text{for } y_i = 1, \quad (3.16)$$

$$x_i^\top w + b \leq -1 + \xi_i \quad \text{for } y_i = -1, \quad (3.17)$$

$$\xi_i \geq 0, \quad (3.18)$$

These three constraints can be expressed by two inequalities which will be used in margin maximization problem:

$$y_i (x_i^\top w + b) \geq 1 - \xi_i \quad (3.19)$$

$$\xi_i \geq 0 \quad (3.20)$$

where the penalization of misclassification is introduced by ξ being the classification error depending on the distance from a misclassified point x_i to the canonical hyperplane that can be seen in Figure 3.2 on the right-hand side. Hence, if the value of ξ is greater than zero, it means that misclassification occurs. As it is seen we allow here for some points to be misclassified but penalize these points appropriately.

With a given training set $\{(x_i, y_i)\}_{i=1}^n$ the goal of penalized margin maximization and data separation can be achieved by solving the following minimization problem of the SVM:

$$\min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \quad (3.21)$$

subject to

$$y_i (x_i^\top w + b) \geq 1 - \xi_i \quad (3.22)$$

$$\xi_i \geq 0. \quad (3.23)$$

where the first term in (3.21) is the inverse margin. Thus, through minimizing this term we maximize the margin. In the second part of the problem, C reflects how good the used algorithm will be in classifying unlabeled future data; in other words, C is the complexity. As the value of C gets smaller, the margin will widen and this parameter provides us with a way to control for the over-fitting problem. However, the smaller C may lead to higher degree of misclassification, as mentioned before. Besides, $\sum_{i=1}^n \xi_i$ serves as an upper bound on the number of training errors.

3 Bankruptcy prediction with Support Vector Machines (SVM)

Now we can proceed to primal Lagrange functional:

$$\min_{w,b,\xi_i} \max_{\alpha_i,\mu_i} L_P = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{y_i (x_i^\top w + b) - 1 + \xi_i\} - \sum_{i=1}^n \mu_i \xi_i, \quad (3.24)$$

where $\alpha_i \geq 0$ and $\mu_i \geq 0$ are Lagrange multipliers.

KKT first order optimality conditions are:

$$\begin{aligned} \frac{\partial L_P}{\partial w_k} = 0 &\iff w_k - \sum_{i=1}^n \alpha_i y_i x_{ik} = 0, \quad k = 1, \dots, d \\ \frac{\partial L_P}{\partial b} = 0 &\iff \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L_P}{\partial \alpha_i} = 0 &\iff y_i (x_i^\top w + b) - 1 + \xi_i \geq 0 \quad i = 1, \dots, n \\ \frac{\partial L_P}{\partial \mu_i} = 0 &\iff \xi_i \geq 0 \quad i = 1, \dots, n \\ &\alpha_i \geq 0 \\ &\mu_i \geq 0 \\ &\alpha_i \{y_i (x_i^\top w + b) - 1 + \xi_i\} = 0 \\ &\mu_i \xi_i = 0 \end{aligned}$$

After we substitute the KKT conditions into the primal problem, we get the dual Lagrangian:

$$\max_{\alpha_i} L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j, \quad (3.25)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \quad (3.26)$$

$$\sum_{i=1}^n \alpha_i y_i = 0. \quad (3.27)$$

So far the hyperplane we are searching for is a linear classifier. However, we know that it is also possible to use SVMs as nonlinear classifiers via the classic kernel trick which enables to embed the data points x_i into a higher dimensional Hilbert space. This trick is possible because of the fact that all the training vectors in the dual problem are expressed as scalar products, $x_i^\top x_j$. This mapping can be shown as:

$$\Phi : \mathbb{R}^d \mapsto \mathbb{H}$$

What we will do is to seek a linear classifier in the feature space \mathbb{H} , which is in fact nonlinear in the original space \mathbb{R} . It is worth stressing that thanks to this kernel trick it is possible to utilize any kernel matrix K defined by $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$ without knowing exactly the high dimensional mapping Φ . In a simple way, instead of using $x_i^\top x_j$ in order to solve the learning problem, we transfer all the data to a higher dimensional space through replacing $x_i^\top x_j$ by $K(x_i, x_j)$. According to Mercer's theorem Martin (1909) any positive semi-definite matrix is a valid kernel matrix meaning that for any dataset x_1, \dots, x_n and any real numbers $\Upsilon_1, \dots, \Upsilon_n$,

the function K should verify

$$\sum_{i=1}^n \sum_{j=1}^n \Upsilon_i \Upsilon_j \geq 0 \quad (3.28)$$

In this work, as mentioned before Gaussian RBF will be applied:

$$K(x_i, x_j) = \exp\{-\|x_i - x_j\| / (2r^2)\} \quad (3.29)$$

where r is the radial bases giving the minimum radius of a data group.

After solving the dual SVM problem, we will find Lagrange multipliers α which take the value of zero for non-support vectors. Support vectors are the training points which are not classified clearly, so that they are either misclassified, or are correctly classified but lie in the margin zone. Accordingly, the α_i values of support vectors will depict non-zero values. As a result, it can be noted that the higher the α_i , the more difficult it is for an observation to be assigned to a class.

After deriving Lagrange multipliers, with the help of the linear relationship we will get the slope of the separating function w :

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3.30)$$

With the acquired values of w , the next step will be the separation of companies into two groups, namely "solvent" and "insolvent" through using the classification rule given by (3.31). However, this is the case when it is possible to classify data linearly in the original space \mathbb{R} .

$$g(x) = \text{sign}(x^\top w + b), \quad (3.31)$$

In (3.31) the threshold parameter is $b = \frac{1}{2} (x_{+1} + x_{-1}) w$. x_{+1} and x_{-1} are two support vectors belonging to different classes lying on the margin boundary. The use of averages over all x_+ and x_- instead of two arbitrarily chosen support vectors is preferable in order to mitigate numerical errors while training the SVM (Moro et al. (2010)). Furthermore, in the linearly separable case, scores of each company can be calculated as:

$$f(x) = x^\top w + b. \quad (3.32)$$

Nevertheless, in this work we are interested in the more general case where data is linearly nonseparable in \mathbb{R} . In such a case, we have a linear classifier $g_N(x)$ in the feature space \mathbb{H} shown as:

$$g_N(x) = \text{sign}\left(\sum_{i=1}^n y_i \alpha_i K(x_i, x_j) + b\right), \quad (3.33)$$

and the scores of a companies are given by:

$$f(x) = \sum_{i=1}^n y_i \alpha_i K(x_i, x_j) + b. \quad (3.34)$$

3 Bankruptcy prediction with Support Vector Machines (SVM)

Next how we can solve this classification problem is important. Due to the fact that the optimization problem in question is a quadratic one it is possible to solve it by the means of quadratic programming (QP); nonetheless, there are significant problems:

1. solving the problem is extremely time consuming
2. it requires a huge memory

To overcome these problems while solving the optimization problem three possible methods can be succinctly introduced. Firstly, the Chunking algorithm developed by Vapnik (1979) can be shown to be an effective method for decreasing the required time and reducing the huge memory requirement while optimizing the SVM in relation to the classical QP optimization. This method uses support vectors having non-zero Lagrange multipliers ($\alpha \neq 0$) to solve the problem and ignores the other observations. Thus, thanks to this method we can maintain a reduced training set and simplify the optimization problem. It is clear that this algorithm depends on the idea that the solution of QP optimization will not be changed if we disregard vectors with $\alpha = 0$.

The steps of the general chunking procedure can be described as:

1. Decompose the original dataset as: training dataset and testing dataset.
2. Optimize the reduced training set by classical QP optimization.
3. Obtain the support vectors by the previous optimization.
4. Test the support vectors in testing dataset.
5. Recombine support vectors and testing errors on a new training set.
6. Repeat from step 2 until errors are reduced enough by 4.

The other method is developed by Osuna and F. (1997) in which the number of observations is not reduced, but instead it is kept constant. The idea is that after solving the QP problem, one of the observations is replaced with the one that contradicts with the KKT optimality conditions.

Both of the mentioned methods utilize the QP solver. Additionally, another method, namely Sequential Minimal Optimization, was introduced by Platt (1998) which solves the relevant optimization problem analytically. In this work, chunking method will be applied.

After solving the dual problem and predicting scores $f(x)$ through learning algorithm, we will find the probability of default (PD) for each company. The PD is a conditional probability which can be defined as the probability of $Y = 1$ given $X = x$. The PD of a company gives the likelihood that the company will not repay its loan, in other words it will default on the loan. Let's define this conditional probability as:

$$\eta(x) = P(Y = 1 | X = x) \quad \forall x \in X, \quad \eta(x) \in [0, 1]$$

In this work, through using least square loss function we get the following formula to calculate PDs from each value of company scores $f(x)$.

$$\hat{\eta}(x) = \frac{f(x_i) + 1}{2} \tag{3.35}$$

All in all, in predicting company failures with the SVM we have two phases, namely training and test phases. During the training phase, the SVM estimates weights w of the financial ratios, and

through doing this it learns the mapping $y = f(x, w)$ that is done by the system. What we look for is an approximating function $f_a(x, w) \sim y$ that can have a good generalization performance on the future data. Following training, we have test phase, where we expect that the output from a machine $o = f_a(x, w)$ is a good estimate of the true response y .

4 Bankruptcy prediction with logistic regression

The other possible method to predict corporate failure by using financial ratios X_j is the logistic model. In this model, it is assumed that $(X_1, Y_1), \dots, (X_n, Y_n)$ is an independently and identically distributed random sample, where $X_j \in \mathbb{R}^d$. The dependent variable Y_j is a Bernoulli random variable taking the value of 1 and 0 referring to bankrupt and non-bankrupt companies in this work, respectively.

Through using logistic regression we will try to fit data to a logit function to predict the PD given some financial ratios $P(Y_j = 1 | X_j = x)$. Because of the fact that the regression is binomial, the generalized linear model (GLM), which was introduced by Nelder and Wedderburn (1972), will be utilized. We can not predict default probability through using a linear regression of $\beta_0 + \beta^\top X$ because here, there is no guarantee that the probability will lie between the values of 0 and 1. Therefore, GLM will be used to predict the probability. The scores are computed by $\beta_0 + \beta^\top X$ meaning that they are linear combinations of the relevant financial ratios. After calculating the scores, the probabilities of default will be estimated with a link function G :

$$\phi(X) = P(Y_j = 1 | X_j = x) = G(\beta_0 + \beta^\top X) \quad (4.1)$$

where $G : \mathbb{R} \rightarrow [0, 1]$ is a known function which is selected as logistic here taking the values between 0 and 1:

$$G(s) = \gamma(s) = \frac{1}{1 + e^{-s}} \quad (4.2)$$

where s gives scores. Hence, scores can also be written as the log of odds:

$$s = \ln \left\{ \frac{\gamma(s)}{1 - \gamma(s)} \right\} \quad (4.3)$$

The real values of β_0, \dots, β_d are not known and required to be estimated. The maximum likelihood procedure will be applied to estimate them. The conditional likelihood function can be written as:

$$L(\beta_0, \dots, \beta_d) = \prod_{j=1}^n [Y_j \gamma(\beta_0 + \beta^\top X_j) + (1 - Y_j) \{1 - \gamma(\beta_0 + \beta^\top X_j)\}] \quad (4.4)$$

As put before, Y_j lies between the values of 0 and 1, so the corresponding conditional log-

4 Bankruptcy prediction with logistic regression

likelihood function can be written as:

$$\log L(\beta_0, \dots, \beta_d) = \sum_{j=1}^n [Y_j \log \gamma(\beta_0 + \beta^\top X_j) + (1 - Y_j) \log \{1 - \gamma(\beta_0 + \beta^\top X_j)\}] \quad (4.5)$$

Then, L or $\log L$ is maximized to get the maximum estimators $\hat{\beta}_0, \dots, \hat{\beta}_d$ of β_0, \dots, β_d and lastly, we obtain the maximum likelihood estimator for the PD (Franke and Hafner (2004)):

$$\hat{\phi}(X) = \gamma(\hat{\beta}_0 + \hat{\beta}_0^\top X) \quad (4.6)$$

Next, it is essential to explain which financial ratios should be added to the model to predict default probabilities in a most accurate way. To select the best model, we will check one of the measures of the goodness of fit, namely Akaike's information criterion (AIC) proposed in Akaike (1974). Through using backward stepwise regression, the analysis will start with all financial ratios and the ratios will be eliminated from the model sequentially. In each step of elimination, the AIC will be calculated. Then, these models are ranked according to their AIC and the model having the lowest AIC will be selected. The AIC can be calculated as:

$$AIC = -2\log L + 2p \quad (4.7)$$

where L is the maximized log likelihood of the model fit and p represents the number of estimated parameters. The model having a good fit to data has a low value of deviance statistic $-2\log L$. Furthermore, $2p$ is added to the AIC to penalize the likelihood for each parameter added to the model because addition of the parameters makes the model more complicated.

After selecting the model with lowest AIC , chi-square test is applied to prove how well the chosen logistic regression model fits the data. To calculate the chi-square statistic, the difference in deviance statistics for the final model and the null model, a model with just an intercept, will be taken. In other words, to get the required test statistic the difference between the maximized value of the likelihood functions for the null model L_N and the final model L_F should be found which can be expressed as:

$$\chi^2 = (-2\log L_N) - (-2\log L_F) \quad (4.8)$$

The degrees of freedom is equal to the number of parameters in the final model.

In the last step, to compare the goodness of fit of logistic and SVM models, the Lorenz curve will be plotted which was introduced by Lorenz (1905). With the help of this graph accuracy of scores s with respect to their predictive power for a company default is visualized for both models. Simply put, the Lorenz curve is a plot of $P(S < s)$ against $P(S < s | Y_j = 1)$. Thus, it is clear that the aim is to understand which of the models classifies more companies as "non-bankrupt" depending on their own dataset, although they went "bankrupt" in reality. And according to the visualization of the level of this misclassification we try to assess the performances of logistic and SVM models and decide which model outperforms the other.

5 Data description

The dataset consists of 20000 solvent and 1000 insolvent German companies. It is obtained from the Credit reform database provided by the Research Data Center (RDC). The analyzed period runs from 1996 to 2002 and it is important to stress that the information regarding insolvent companies is acquired 2 years before the insolvency occurred. The label variable y takes the value of -1 for solvent companies and 1 for the last report of a company before its insolvency.

We have 28 financial variables belonging to the companies. The variables are such as cash, earnings before interest and taxes (EBIT), amortization and depreciation (AD), intangible assets (ITGA) and lands and buildings (LB). Through using these variables 24 financial ratios are computed which are put into four groups so called profitability, leverage, liquidity, and activity shown in Table 5.1.

5.1 Data cleaning and model selection

The SVM is sensitive to outliers, because outlier points play the most important role in determining the decision hyperplane. To enhance the SVM the effect of outliers are decreased through equalizing the data points greater than the upper outlier to this upper value and similarly replacing the values smaller than the lower outlier by this lower value. Upper and lower limits are calculated by $Q75 + 1.5 * IQ$ and $Q25 - 1.5 * IQ$, respectively where Q represents the quantile. Besides, it is required to note that in the year of 1996 we have just insolvent companies so that the data of this year will not be incorporated to the calculations. Descriptive statistics of financial ratios are given in Table 5.2.

The next and very important step is to select the most powerful financial ratios in separating the companies into two groups; solvent and insolvent. In this work, backward and forward stepwise selection techniques are utilized in MATLAB 7.0 and then the results are compared to show how important it is to decide which financial ratios should be incorporated to the SVM model. Due to the fact that we have too many financial ratios, the critical value is taken as 0.01 in both techniques to receive the most important ones and eliminate others.


In the case of backward stepwise selection, the procedure starts with all variables, and the variables having less significance according to the chosen critical level are then sequentially removed. This procedure will continue until all remaining variables are statistically significant. After the application of this method, the model comprising $x_1, x_3, x_5, x_6, x_8, x_{11}, x_{13}, x_{18}$ and x_{24} is chosen. The selection of the financial ratios poses a meaningful picture, since with this selection the model has at least one financial ratio from each category seen in Table 5.3.

At this point, it is required to mention the importance of the selected financial ratios. The return on assets defines how profitable a company's assets are in terms of generating revenues. Thus, it shows how efficiently a company can convert its assets into net income. Therefore, using this ratio in separation of companies seems highly reasonable. Additionally, other chosen


Ratio No.	Definition	Ratio	Category
x1	NI/TA	Return on assets	Profitability
x2	NI/Sales	Net profit margin	Profitability
x3	OI/TA	Operating Income/Total assets	Profitability
x4	OI/Sales	Operating profit margin	Profitability
x5	EBIT/TA	EBIT/Total assets	Profitability
x6	(EBIT+AD)/TA	EBITDA	Profitability
x7	EBIT/Sales	EBIT/Sales	Profitability
x8	Equity/TA	Own funds ratio (simple)	Leverage
x9	(Equity-ITGA)/ (TA-ITGA-Cash-LB)	Own funds ratio (adjusted)	Leverage
x10	CL/TA	Current liabilities/Total assets	Leverage
x11	(CL-Cash)/TA	Net indebtedness	Leverage
x12	TL/TA	Total liabilities/Total assets	Leverage
x13	Debt/TA	Debt ratio	Leverage
x14	EBIT/Interest expenses	Interest coverage ratio	Leverage
x15	Cash/TA	Cash/Total assets	Liquidity
x16	Cash/CL	Cash ratio	Liquidity
x17	CA/CL	Current ratio	Liquidity
x18	WC/TA	Working Capital	Liquidity
x19	CL/TL	Current liabilities/Total liabilities	Liquidity
x20	Sales/TA	Asset turnover	Activity
x21	Sales/INV	Inventory turnover	Activity
x22	Sales/AR	Account receivable turnover	Activity
x23	Purchases/AP	Account payable turnover	Activity
x24	Log(TA)	Log(Total assets)	Activity

Table 5.1: Definitions of financial ratios.

Ratio	Min	Median	Max	IQR	Sta.Dev.
x_1	0.00	0.01	0.16	0.05	0.05
x_2	0.00	0.01	0.10	0.03	0.03
x_3	0.00	0.03	0.26	0.08	0.07
x_4	0.00	0.02	0.16	0.05	0.05
x_5	0.01	0.05	0.27	0.09	0.07
x_6	0.05	0.10	0.42	0.12	0.09
x_7	0.01	0.05	0.27	0.09	0.07
x_8	0.05	0.16	0.91	0.28	0.20
x_9	0.02	0.18	1.69	0.55	0.54
x_{10}	0.12	0.21	0.82	0.23	0.21
x_{11}	0.11	0.31	1.40	0.43	0.24
x_{12}	0.34	0.60	1.72	0.45	0.22
x_{13}	0.01	0.15	0.97	0.32	0.20
x_{14}	0.45	1.92	17.23	5.56	5.75
x_{15}	0.01	0.14	2.16	0.71	0.75
x_{16}	0.01	0.08	0.96	0.32	0.31
x_{17}	1.04	1.54	5.73	1.55	1.44
x_{18}	0.02	0.20	0.98	0.39	0.23
x_{19}	0.54	0.86	1.18	0.45	0.19
x_{20}	0.34	0.57	2.61	0.75	0.73
x_{21}	0.02	0.09	0.53	0.17	0.14
x_{22}	0.05	0.09	0.34	0.10	0.07
x_{23}	0.03	0.07	0.29	0.08	0.07
x_{24}	6.22	6.82	9.99	1.25	0.77

Table 5.2: Descriptive statistics for financial ratios. IQR is the interquartile range.  Descriptive

Ratio No.	Definition	Ratio	Category
x_1	NI/TA	Return on assets	Profitability
x_3	OI/TA	Operating Income/Total assets	Profitability
x_5	EBIT/TA	EBIT/Total assets	Profitability
x_6	(EBIT+AD)/TA	EBITDA	Profitability
x_8	Equity/TA	Own funds ratio (simple)	Leverage
x_{11}	(CL-Cash)/TA	Net indebtedness	Leverage
x_{13}	Debt/TA	Debt ratio	Leverage
x_{18}	WC/TA	Working Capital	Liquidity
x_{24}	Log(TA)	Log(Total assets)	Activity

Table 5.3: The selected financial ratios via backward stepwise method to be used in the SVM model.  SVMbackward

5 Data description

profitability financial ratios, EBIT/Sales and (EBIT+DA)/TA, are again the measures of the efficiency of a company in generating returns from its assets. In the calculation of second financial ratio depreciation and amortization are added to EBIT, since they are not counted as operating expenses. The other significant financial ratio in the same category is operating profit margin which reflects the strength of the company in paying its fixed costs, such as interest and debt. All in all, it can be said that the higher these profitability ratios, the lower is the company's financial risk. As to the leverage category, own funds ratio (simple), net indebtedness and debt ratio are selected via backward stepwise selection. Own funds ratio tells us the solvency level of a company through giving its ability to raise money to remain viable. Net indebtedness reflects a company's debt situation after subtracting its cash from the value of its liabilities. Debt ratio provides similar information. Through looking at debt ratio it is possible to grasp the ability of a company to secure its financing. Hence, the selected leverage financial ratios are important variables for assessing the ability of a company to sustain risk. In liquidity category the ratio of working capital to total assets is chosen where working capital is given by current assets; thus, the relevant ratio gives the amount of operating capital of the company that can be used for growth or acquisitions. When this ratio has a positive sign, it means that the company has adequate funds to meet its operational expenses and short-term debt. Lastly, for the category of activity, log of total assets will be used which is the log of the size of the company. A company's size can be seen as a risk factor for bankruptcy. It is mostly believed that large companies are unlikely to go bankrupt. However, there are some instances of bankrupted huge enterprises in real life that contradict this claim, such as Enron, Barings Bank, and Worldcom. Hence, we can not have an exact comment on the effect of company size on the company default risk.

As to the other technique of variable selection, forward stepwise selection, the procedure starts with an intercept model and then proceeds through adding financial ratios sequentially to the model and stop when no additional financial ratio can improve the fit significantly. According to this method x_1 , x_3 , x_5 , x_6 , x_8 , x_{19} and x_{24} are chosen, and again we have at least one variable relating to each category of financial ratios as represented in Table 5.4. Compared to backward stepwise selection method we have fewer amounts of variables, but six of them are common after using both techniques. Here, the only different financial ratio from those selected by backward technique is the ratio of current liabilities to total liabilities which is in the liquidity category. Current liabilities are all debts or obligations of the company due within twelve months; thus, it is an important measure which provides information regarding the company's liquidity position. Hence, this ratio can be seen as a crucial measure reflecting the company's debt structure and so gives an aspect of a company's financial performance as well. All in all, including this ratio to the SVM model seems reasonable.

Ratio No.	Definition	Ratio	Category
x_1	NI/TA	Return on assets	Profitability
x_3	OI/TA	Operating Income/Total assets	Profitability
x_5	EBIT/TA	EBIT/Total assets	Profitability
x_6	(EBIT+AD)/TA	EBITDA	Profitability
x_8	Equity/TA	Own funds ratio (simple)	Leverage
x_{19}	CL/TL	Current liabilities/Total liabilities	Liquidity
x_{24}	Log(TA)	Log(Total assets)	Activity

Table 5.4: The selected financial ratios via forward stepwise method to be used in the SVM model.  SVMforward

6 Empirical Results

6.1 Results on the SVM model

In this part, the aim is to calculate the score values, then the PDs for each company and lastly, to obtain the classification of two groups of bankrupt and non-bankrupt companies. It is known that after the selection of variables the result of the margin maximization problem which gives the effectiveness of the SVM in classification depends heavily on the values of C , the capacity, and r , the kernel's parameter. As mentioned before, r is the minimum radius containing the data. Additionally, the selection of kernel is critically important in determining how good the separation can be. As stated earlier, Gaussian RBF is applied in this work. The optimization problem of the SVM model is solved with MATLAB 7.0.

For the model with the 9 financial ratios found by using backward stepwise selection, a training set and test set are taken in a 2 : 1 ratio. In the training set there are randomly selected 100 solvent and 100 insolvent companies, whereas in the test set we have 50 companies for each group. It is observed that increasing the number of companies in the training set worsens the performance of the SVM model as a classifier. Through changing the values of C and r we get different performances given by the accuracy ratio (AR). At this point, it is crucial to find the best combination of C and r which gives the highest AR. To achieve this result grid-search with exponentially growing sequences of these two relevant parameters is applied, where $C \in \{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}$ and $r \in \{2^{-15}, 2^{-13}, \dots, 2^1, 2^3\}$. Through using cross validation each possible pair of parameters is checked and then the parameters giving the highest cross validation accuracy are selected. The highest AR of 72% is obtained by $C = 16$ and $r = 2$. The AR results of other combinations of the parameters can be seen in Table 6.1. Next, we check what will happen when included information is decreased. For example, net indebtedness ratio is dropped and after applying cross validation to the model with 8 financial ratios the highest AR of 75% is achieved with $C = 128$ and $r = 4$, the visualization of which can be seen in Figure 6.1 derived in R 2.12.0. As mentioned before, we found the PDs from company scores through using least square loss function and because of the fact that what we found is just the estimates of the probabilities some values are outside the interval $\{0, 1\}$. To solve this problem the values higher than one are set equal to one and the values smaller than zero are set equal to zero. To conclude, although a statistical method is used to select the best financial ratios, we see that dropping a financial ratio can improve the model's performance. Thus, incorporating too much information to the model does not necessarily lead to improvement of the model. It is obvious that it is not easy to find the model with the highest AR.

As to the model with 7 financial ratios based on forward stepwise selection, we use the same training and test sets and we apply cross validation method depending on the same intervals used previously for the parameters of C and r . In this case, AR posted the best value at 64% with $C = 256$ and $r = 2$. It is observed that using different financial ratios based on different variable selection procedures affects the performance of the SVM model significantly. How the

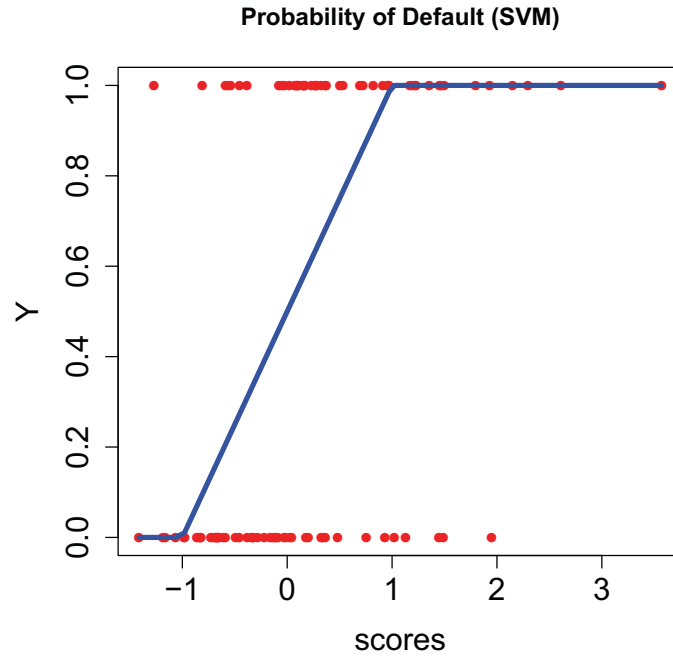



Figure 6.1: The SVM model with AR=75% where the blue line represents the PDs of the companies and red dots are the scores. [SVMpd](#)

C	r	AR
16.00	0.06	49%
0.03	8.00	50%
0.50	4.00	53%
0.25	1.00	55%
0.25	0.25	57%
16.00	0.25	58%
2.00	1.00	59%
0.50	0.50	60%
4.00	1.00	65%
8.00	1.00	68%

Table 6.1: The effect of different pairs of C and r on AR of the SVM model composed of the financial ratios given by backward selection method. [SVMcrbackward](#)

C	r	AR
128.00	0.25	43%
0.03	0.50	50%
8.00	0.13	51%
0.50	0.25	52%
2.00	16.00	54%
4.00	0.50	56%
16.00	0.50	58%
256.00	1.00	61%
0.06	0.03	63%

Table 6.2: The effect of different pairs of C and r on AR of the SVM model composed of the financial ratios given by forward selection method.  SVMcrforward

different combinations of parameters influence the values of AR can be observed in Table 6.2.

To visualize how the effectiveness of the SVM classification varies according to the parameters of C and r , different instances will be analyzed in two dimensions through using the model with x_1 and x_3 consisting of randomly chosen 100 solvent 100 insolvent companies. The graphs are produced with R 2.12.0. As a reminder, x_1 and x_3 refer to return on assets and the ratio of operating income to total assets, respectively both of which belong to the profitability group. Intonation of the related figures depends on the level of scores f . As it is known, the higher the scores, the higher is the PD of the company. When the score value gets larger, the color should become blue; in other words, the more blue color represents the higher score value and in parallel the higher PD. As to the areas with companies having low or negative scores, the color should get more red. Thus, the regions of the graph filled with successful companies are expected to be red in the case of a good classification. Additionally, triangles and circles are solvent and insolvent companies, respectively and the ones which are filled with black color represent the support vectors.

Firstly, we will check how the performance of the SVM changes according to the different values of r . In Figure 6.2 the values of the parameters are $C = 100$ and $r = 1$. In this case, it can obviously be seen that there is no successful classification. The reason might be that C remains too high according to the value of r . Then, we increase the value of r to 8 while keeping the value of C invariant and SVM starts to classify the companies given in Figure 6.3. As it is seen, blue areas are filled with insolvent companies and red ones with solvent companies. Next, r is increased to 100 while keeping C constant presented in Figure 6.4 so as to figure out what happens when r takes a very high value. It can be observed that the colorful areas become very small and unclear meaning that we increase the minimum radius r to an undesirable high level. There are some correct classifications on the left bottom, though.

It is already stated that the capacity C has an inverse relationship with the width of the margin. The analysis of the impact of the changes in C on the classification performance of SVM will be performed while keeping the value of r at 8 in each case. Firstly, the value of C is taken as 0.1 in Figure 6.5. As it is seen, almost all of the companies become support vectors, although SVM recognizes the few successful companies on the left bottom side of the figure. It is obvious that the value of C needs to be raised. In Figure 6.6, we increase the value of C slightly to 0.5 to observe how sensitive the SVM is to the capacity parameter, and the result changes

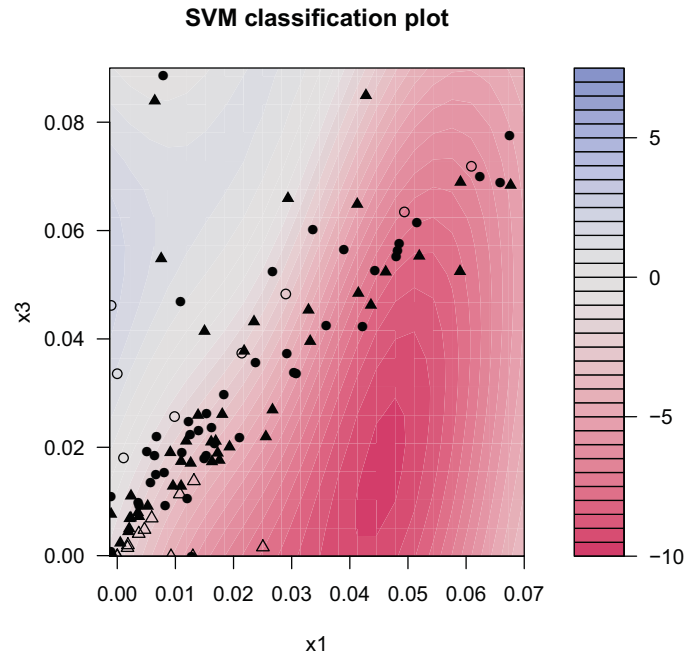


Figure 6.2: Assessment of the SVM classification performance in two dimensions where $C = 100$ and $r = 1$. ●○▲△SVMc100r1

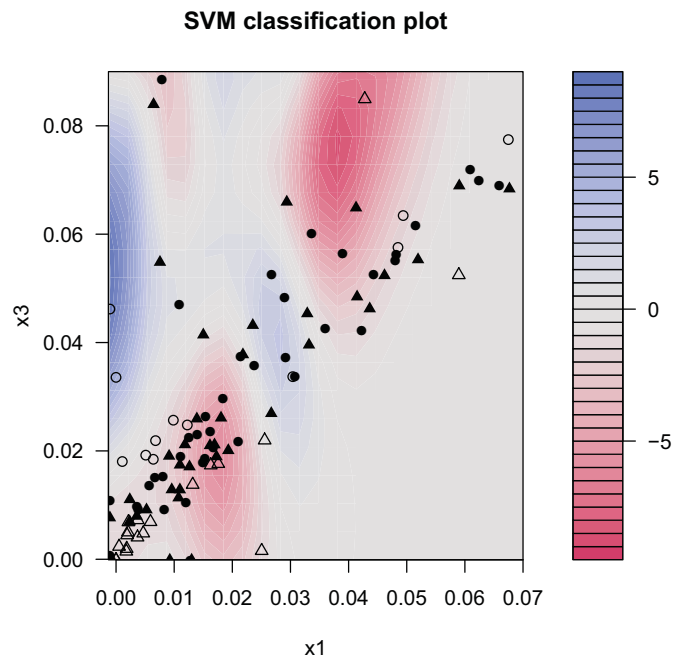


Figure 6.3: Assessment of the SVM classification performance in two dimensions where $C = 100$ and $r = 8$. ●○▲△SVMc100r8

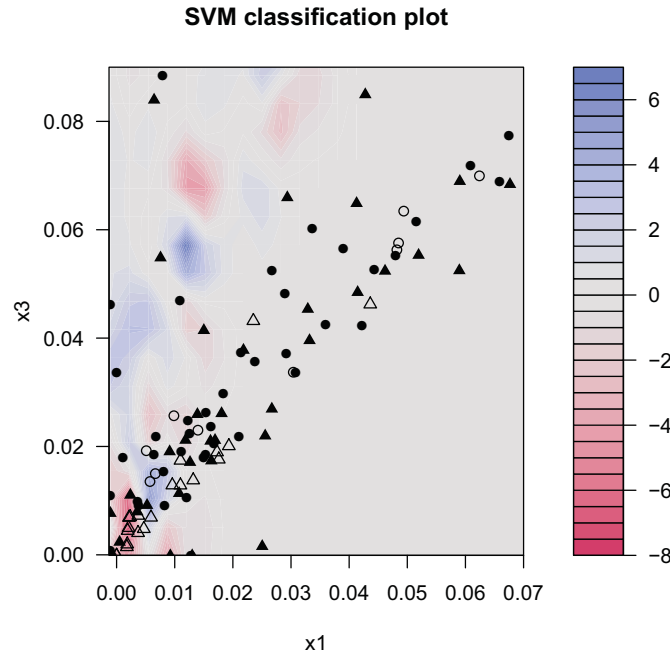


Figure 6.4: Assessment of the SVM classification performance in two dimensions where $C = 100$ and $r = 100$. $\text{SVM}_{C=100, r=100}$

considerably with blue and red clusters being full with correct companies and the number of support vectors decreases as well. Next, in Figure 6.7, the parameters are $C = 170$ and $r = 8$ and it is observed that the groups are separated with blue and red regions, whereas Figure 6.8 derived with the parameters of $C = 10000$ and $r = 8$ indicates that with very high values of C the distance between two groups of companies gets very narrow that SVM starts not to recognize the bankrupt companies yielding unsuccessful classification.

After analyzing the crucial effect of the selection of C and r on the SVM's performance, it becomes clearer that a statistical method should be applied in order to find the optimum pair of parameters giving the best SVM classification on data.

6.2 Results on logistic regression

In this section, the results of logistic regression model will be given. The data includes 500 solvent and 500 insolvent companies randomly chosen from the same dataset provided by RDC. As mentioned before, the response variable $Y \in \{0, 1\}$ is binary and the value of -1 used before for the solvent companies is changed with 0 here and we leave the value of 1 for the insolvent companies unchanged.

To choose the best model with the given financial ratios the AIC stepwise algorithm is applied. We adopt backward elimination, where the procedure starts with all financial ratios. Then, financial ratios are removed sequentially and the AIC is computed at each step. As put before, among the candidate models, the model with the lowest AIC is desirable. The AIC touched the lowest value at 1164.92 which gives the final model composed of $x_1, x_3, x_{12}, x_{15}, x_{16}, x_{18}$,

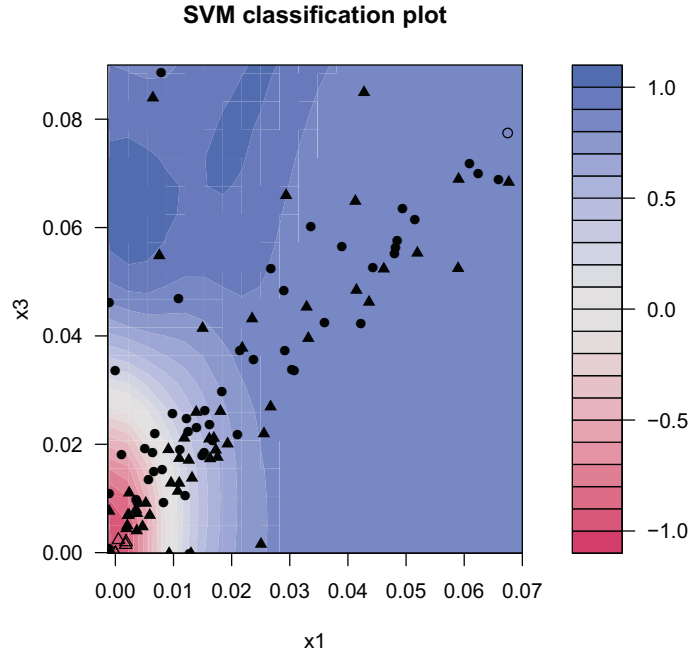


Figure 6.5: Assessment of the SVM classification performance in two dimensions where $C = 0.1$ and $r = 8$. SVMc01r8

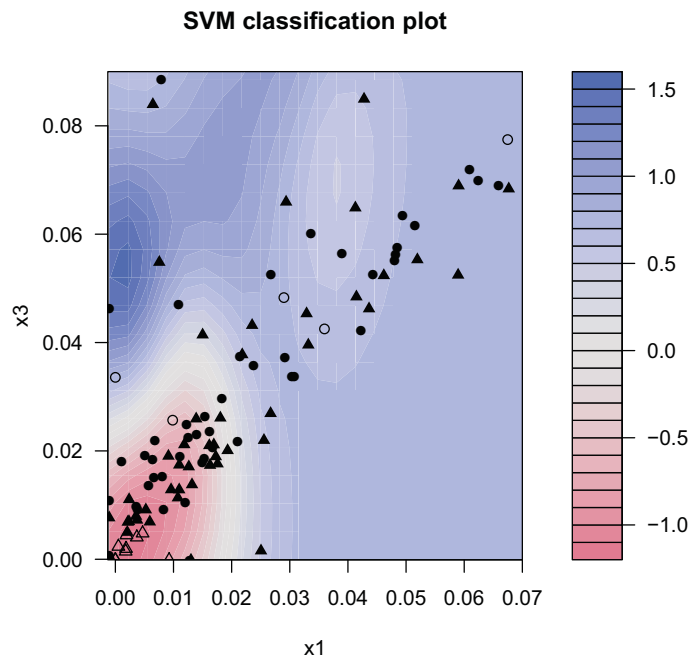


Figure 6.6: Assessment of the SVM classification performance in two dimensions where $C = 0.5$ and $r = 8$. SVMc05r8

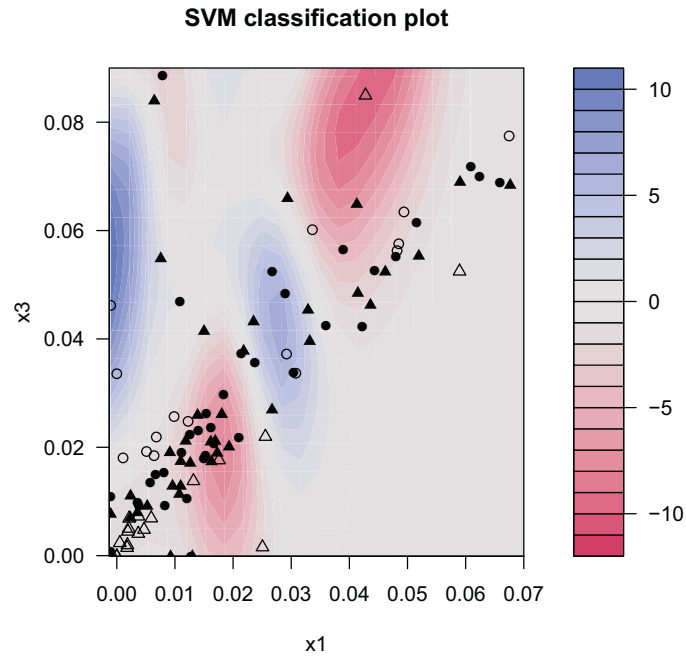


Figure 6.7: Assessment of the SVM classification performance in two dimensions where $C = 170$ and $r = 8$. SVMc170r8

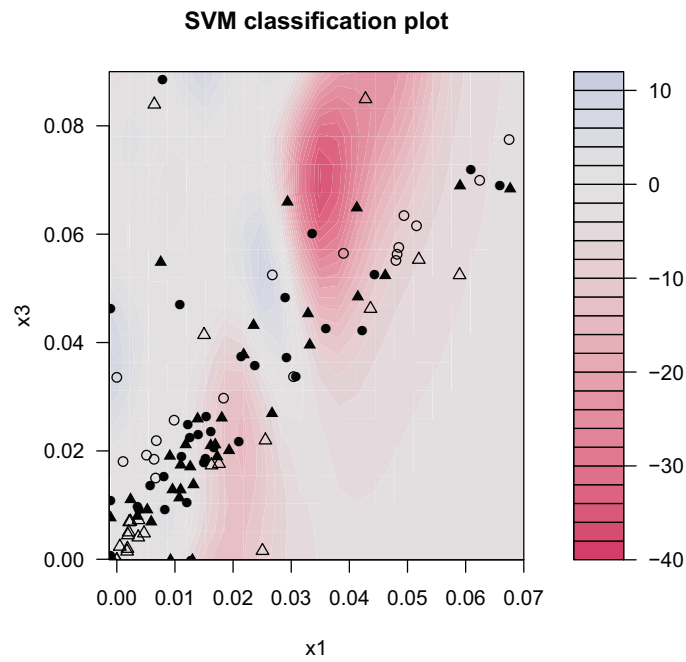



Figure 6.8: Assessment of the SVM classification performance in two dimensions where $C = 10000$ and $r = 8$. SVMc10000r8

Ratio No.	Definition	Ratio	Category
x_1	NI/TA	Return on assets	Profitability
x_3	OI/TA	Operating Income/Total assets	Profitability
x_{12}	TL/TA	Total liabilities/Total assets	Leverage
x_{15}	Cash/TA	Total liabilities/Total assets	Liquidity
x_{16}	Cash/CL	Cash ratio	Liquidity
x_{18}	WC/TA	Working Capital	Liquidity
x_{20}	WC/TA	Asset turnover	Activity
x_{21}	Sales/INV	Inventory turnover	Activity
x_{22}	Sales/AR	Account receivable turnover	Activity
x_{23}	Sales/AP	Account payable turnover	Activity
x_{24}	Log(TA)	Log(Total assets)	Activity

Table 6.3: The selected financial ratios via AIC to be used in the logistic regression.  GLMresults

x_{20} , x_{21} , x_{22} , x_{23} and x_{24} . We have at least one variable from each category of financial ratios given in Table 6.3.

In Table 6.3 there are some variables not explained before. In the leverage group, we have the ratio of total liabilities to total assets (x_{12}) which reflects the company's total level of liabilities and its capacity of borrowing; hence, this ratio can be seen as a long-term solvency ratio. As to the liquidity category, we have one new financial ratio, namely cash ratio (x_{16}) which is a crucial measure of the company's ability in repaying its short-term debt. Moreover, in the activity part, there are three financial ratios needed to be introduced, which are inventory turnover (x_{21}), account receivable turnover (x_{22}) and account payable turnover (x_{23}). x_{21} is calculated through dividing sales by inventory that gives how many times the company's inventory is sold and used over a period. A low level of this ratio indicates unfavorable amount of sales. x_{22} shows how successful the company is in collecting sales on credit during the year. When this ratio is high, it means that the company has effective credit policies that it can successfully convert its accounts receivable into cash. As to the x_{23} , it provides information regarding how many times per period the company pays off its suppliers. Similar to x_{22} , when this ratio takes a high value, it gives positive signal about the company in paying its average payable amount.

As given in a previous section, we will use GLM in order to estimate the PD. Firstly, we calculate the scores by $\beta_0 + \beta^\top X$ and then switch to computation of the PD $\phi(x)$:

$$\phi(X) = P(Y_j = 1 | X_j = x) = G(\beta_0 + \beta^\top X)$$

The results of the model with chosen financial ratios is summarized in Table 6.4 and the visualization of the model created with R 2.12.0 can be seen in Figure 6.9.

In order to test the goodness of fit we compare the null model, a model with just an intercept, with our final model whose results are given in Table 6.4. The chi-square statistic is calculated by the difference between the deviances of the null and final models which is $1386.3 - 1140.9 = 245.4$ and the degrees of freedom is the differences in degrees of freedom of the final and null models being 11. It is required to say that for the deviance of the final model we use residual deviance. As a result, p-value of zero depending on the calculated high chi-square statistic

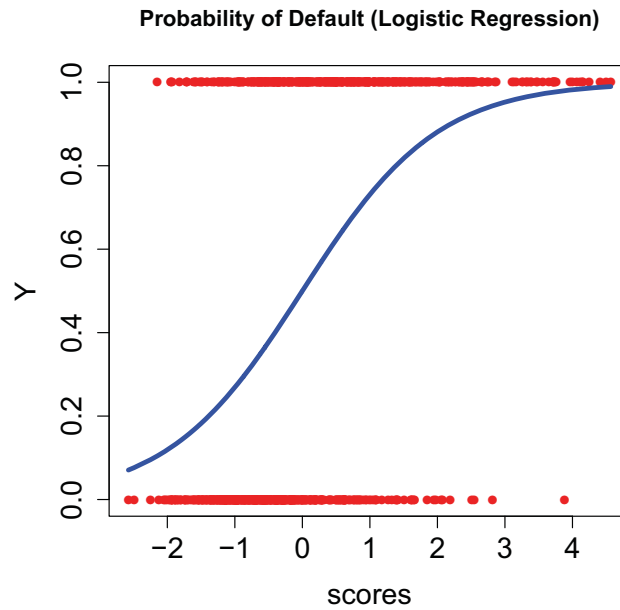


Figure 6.9: The logistic model where the blue line represents PDs of the companies and red dots are the scores. [LOGISTICpd](#)

	Estimate	Sta. Dev.	z value	P(> z)
(Intercept)	-4.30	1.31	-3.28	0.00**
x1	-7.09	3.75	-1.89	0.06*
x3	14.90	2.96	5.03	0.00***
x12	1.07	0.48	2.24	0.02**
x15	0.39	0.13	3.00	0.00***
x16	0.85	0.47	1.79	0.07*
x18	1.13	0.52	2.19	0.03**
x20	0.56	0.16	3.58	0.00***
x21	1.59	0.62	-2.57	0.01**
x22	-2.37	1.00	-2.36	0.02**
x23	-6.38	1.84	-5.39	0.00***
x24	0.54	0.18	3.00	0.00**
Overall model fit				
Null model -2 log likelihood	1386.30			
Full model -2 log likelihood	1140.90			
Chi-square	245.40			
Degrees of freedom	11.00			

Table 6.4: GLM results and overall model fit.*** indicates significance at 1% level, ** indicates significance at 5% level, * indicates significance at 10% level. [GLMresults](#)

shows that the final model fits significantly better than the null model.

As it is seen in Table 6.4, in the final model all the coefficients of financial ratios are statistically significant. Nonetheless, it is important to comment on the sign of the coefficients of the financial ratios and figure out whether they provide a reasonable result. It is found that the higher the return on assets, the lower PD the company has which seems highly sensible. A high value of return on assets usually means that the company can successfully generate income through its investments. On the other hand, the coefficient of the other profitability ratio of operating income to total assets has a positive sign. This means that the higher this ratio, the higher is company's default risk. However, in reality a high value of this ratio indicates the high efficiency level of the company in its operations giving rise to lower financial risk. Hence, the coefficient of this ratio seems implausible. As to the leverage part, the regression result implies that the ratio of total liabilities to total assets has a positive effect on the PD confirming the fact that the higher the amount of total liabilities with respect to total assets, the more likely a company will default on a loan. The other positive signs are obtained for the liquidity ratios of cash to total assets, cash ratio and current ratio implying that high values of these ratios put the company in a risky position. These results can be seen as unreasonable. Nonetheless, a very high level of these ratios might reflect the problem that the company does not use its funds to generate higher returns through profitable investments. Thus, in the long run the company may face an unfavorable financial situation without an effective funds management. Next, according to the results, increases in the asset turnover ratio raise the company's default risk, but it is not the case in reality, because a high value of this ratio proves that the company can efficiently generate money in sales for every dollar of assets which makes it stronger in financial terms, not weaker. Moreover, we found that the higher the activity financial ratios of inventory turnover, accounts receivable turnover and accounts payable turnover, the less the PD the company has. As put before, inventory turnover ratio provides information about how many times the company's inventory is sold and replaced in a specific period of time. When this ratio is low, it means that the company suffers from a weak performance in sales; thus, the negative sign of its coefficient can make sense. A high accounts receivable and payable turnover ratios can be seen as a positive signals for the company's financial situation which is in parallel to the results of the logistic regression. A high value of accounts receivable turnover ratio means that the company has successful credit policies and can efficiently turn its accounts receivables into cash. Similarly, a high value of accounts payable ratio tells us that the company is able to repay its debt to its suppliers with no problems. Lastly, as mentioned before, it is widely believed that large companies have low default risk. Nevertheless, according to the final model the greater the company size, the higher is the PD. The reason of why we found the signs of the most coefficients to be incompatible with the economic theory might be the non-linear relationship of financial ratios with scores which shows the poor performance of the logistic regression in estimating the companies' PDs.

6.3 Comparison of the SVM and logistic models

After presenting the results of the logistic regression, the comparison of its performance with that of the SVM will be provided with the help of a Lorenz curve obtained in R 2.12.0.

Firstly, the estimation results of both models are given in Tables 6.5 and 6.6. As it is seen, in predicting the bankrupt companies, the SVM delivers a better performance than the logistic

	Bankrupt (estimated)	Non-bankrupt(estimated)
Bankrupt(data)	40 (80%)	10 (20%)
Non-bankrupt (data)	15 (30%)	35 (70%)

Table 6.5: Accuracy of the SVM model.

	Bankrupt (estimated)	Non-bankrupt(estimated)
Bankrupt(data)	329 (65.80%)	171 (34.20%)
Non-bankrupt (data)	111 (22.20%)	389 (77.80%)

Table 6.6: Accuracy of the GLM results.

model, i.e. the type II error is smaller in the case of the SVM model. On the other hand, this picture is reversed when estimating the non-bankrupt companies. Here the type I error of the SVM model has a greater value compared to that of the logistic model. However, when we look at the overall performances of the models, the SVM model has AR of 75%, whereas logistic one has a lower value of 71.8%. To see the performance difference more clearly we will check the Lorenz curves of both models.

Lorenz curves are presented in Figure 6.10 where the better performance of the SVM over logistic model can be analyzed, since in the most parts of the graph the blue line lies under the red one. This result is expected, because SVM is a non-parametric classification method which does not require an extremely strict assumption of specific parametric form of the model in contrast to the linear logistic regression in order to predict the PDs. As it is known, we are using financial ratios to estimate scores and then PDs and it is very likely to have a non-linear relationship between scores and financial ratios contradicting the logistic model's assumption of linear dependence.

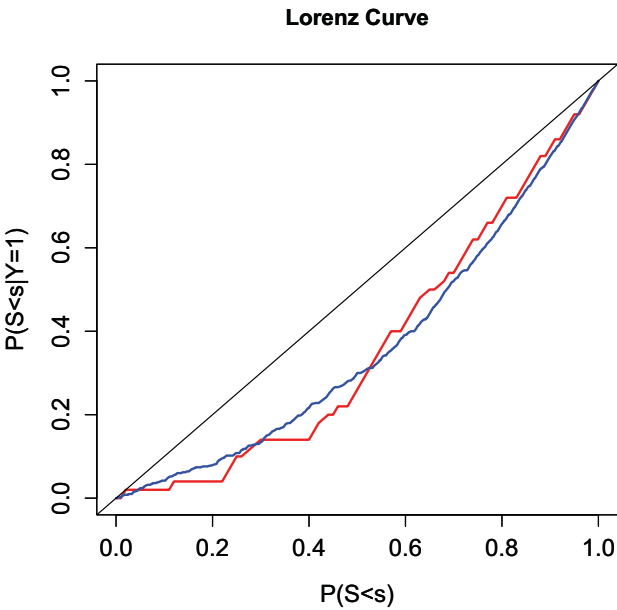



Figure 6.10: Lorenz curves of SVM (blue) and logistic models (red).  LORENZsvm and LORENZlogistic

7 Conclusion

In this work, we applied SVM in order to predict company failure based on financial ratios and compared its prediction results with those of logistic regression model. To achieve the best prediction power of the SVM model we utilized a grid-search technique so as to select the optimal values of radius r used in kernel function and capacity C . In the best case overall testing accuracy ratio obtained through applying the SVM is found at 75%; on the other hand, this ratio reaches the peak at 71.8% via the logistic model. Through Lorenz curves we visualize the result that the SVM model poses a stronger company bankruptcy prediction performance over the logistic model. As we go into more detail, the SVM outperforms the logistic model in detecting the insolvent companies. However, in predicting solvent companies the better performance of the logistic model is demonstrated.

The performance of the SVM model depends highly on various factors, namely sizes of training and test sets, the ratio of training set to test set, the selection of C and r , the utilization of the proper kernel function and the selection of input variables which are financial ratios in our work. Through changing these factors it is possible to improve the classification power of the SVM model. Furthermore, it is required to mention that the SVM has an important limitation regarding the size of the training set. The larger this set, the worse the SVM results get. All in all, the SVM can be improved with further research in an effort to make it more compatible with large datasets.

Based on our favorable results regarding the SVM model we can conclude that for financial institutions the SVM model can be seen as a robust alternative in building bankruptcy prediction models.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–722.
- Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, pages 589–609.
- Altman, E., R., H., and Narayanan, P. (1977). Zeta analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, 1:29–54.
- Aziz, A., Emanuel, D., and Lawson, G. H. (1988). Bankruptcy prediction - an investigation of cash flow based models. *Journal of Management Studies*, 25(5):419–437.
- Beaver, W. (1966). Financial ratios as predictors of failures. *Journal of Accounting Research*, 5:71–111.
- Cortes, C. V. V., editor (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Dimitras, R., Zopounidis, C., Slowinski, A., and Susmaga, R. (1999). Business failure prediction using rough sets. *European Journal of Operational Research*, 114:263–280.
- Duffie, D., Wang, K., and Saita, L. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics, Elsevier*, 3:635–665.
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis, Supplement: Outlook for the Securities Industry*, 7(2):1477–1493.
- Fan, A. and M., P. (2000). Selecting bankruptcy predictors using a support vector machine approach. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, pages 354–359.
- Fitzpatrick, P. (1932). A comparison of the ratios of successful industrial enterprises with those of failed companies. *Journal of Accounting Research*.
- Franke, J. H. W. and Hafner, M. C., editors (2004). *Statistics of Financial Markets*. Springer-Verlag, Berlin Heidelberg.
- Frydman, H., Kao, D. L., and Altman, E. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*, 40(1):269–291.
- Gale, D. K. H. W. and Tucker, A. W., editors (1951). *Linear Programming and the Theory of Games, in Activity Analysis of Production and Allocation, T. C. Koopmans (ed.)*. John Wiley and Sons, New York.

Bibliography

- Glennon, D. and Nigro, P. (2005). Measuring the default risk of small business loans: A survival analysis approach. *Journal of Money, Credit and Banking*, Blackwell Publishing, 37(5):923–47.
- Hastie, T. T. R. and Friedman, J. H., editors (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
- Lorenz, M. O. (1905). Method for measuring concentration of wealth. *JASA*, (9):209–219.
- Mangasarian, K. P. and Bennett, O. L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, pages 1:23–24.
- Martin, D. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character*, 209:415–446.
- Martin, D. (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance*, pages 249–276.
- Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.
- Min, J. H. and Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4):603–614.
- Moro, R., Hoffmann, L., and Härdle, W. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131.
- Moro, R., Hoffmann, L., and Härdle, W. (2010). Learning machines supporting bankruptcy prediction. *SFB 649 "Economic Risk"*.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, series A*, Blackwell Publishing, 135(3):370–384.
- Osuna, E. F. R. and F., G., editors (1997). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. IEEE Press, New York.
- Platt, J. C. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Technical Report*, MSR-TR-98-14:109–131.
- Ramser, J. and Foster, L. (1931). A demonstration of ratio analysis. *Bureau of Business Research*, Bulletin no. 40.
- Shin, K. S., Kim, H. J., and Lee, T. S. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1):127–135.
- Shumway, T. and Hope, R. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74:101–124.
- Sierra, L. and Bechetti, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking and Finance*, pages 2099–2120.

- Tam, K. and Kiang, M. (1992). Managerial application of neural networks: the case of bank failure prediction. *Management Science*, 38(7):926–947.
- Tikhonov, A. N. and Arsenin, V. Y., editors (1997). *Solution of Ill-posed Problems*. W. H. Winston, Washington, DC.
- Vapnik, K. and Chervonenkis, A. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280.
- Vapnik, K. and Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, 16(2).
- Vapnik, V., editor (1979). *Estimation of Dependencies Based on Empirical Data*. Springer, Nauka, Moscow.
- Wang, L., editor (2005). *Support Vector Machines: Theory and Applications*. Springer-Verlag, Berlin Heidelberg.
- Wiginton, J. (1980). A note on the comparison of logit and discriminant models of consumer credit behaviour. *Journal of Financial and Quantitative Analysis*, 15(3):757–770.
- Wilcox, A. (1971). A simple theory of financial ratios as predictors of failure. *Journal of Accounting Research*, pages 389–395.
- Wilson, N., Hope, R., and Summers, B. (1999). Predicting corporate failure and payment behaviour: Addressing some pertinent issues for practitioners. *Credit Management Research Centre, Leeds University Business School*.
- Zavgren, C. (1983). The prediction of corporate failure: The state of the art. *Journal of Accounting Literature*, 2:1–38.
- Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 20(0):59–82.

List of Figures

3.1	The minima f_{opt} and \hat{f}_n of the expected (R) and empirical (\hat{R}) risk functions. . .	11
3.2	The margin in a linearly separable and non-separable case.	12
6.1	The SVM model with AR=75%.	30
6.2	SVM classification performance ($C = 10000, r = 1$).	32
6.3	SVM classification performance ($C = 100, r = 8$).	32
6.4	SVM classification performance ($C = 100, r = 100$).	33
6.5	SVM classification performance ($C = 0.1, r = 8$).	34
6.6	SVM classification performance ($C = 0.5, r = 8$).	34
6.7	SVM classification performance ($C = 170, r = 8$).	35
6.8	SVM classification performance ($C = 10000, r = 8$).	35
6.9	The logistic model.	37
6.10	Lorenz curves.	40

List of Tables

2.1	PDs used in rating structure finance securities.	6
2.2	Rating grades and capital requirements.	7
5.1	Definitions of financial ratios.	24
5.2	Descriptive statistics	25
5.3	Financial ratios chosen via backward stepwise method	25
5.4	Financial ratios chosen via backward stepwise method	27
6.1	The effect of different pairs of C and r on AR (backward selection method) . .	30
6.2	The effect of different pairs of C and r on AR (forward selection method). . .	31
6.3	The selected financial ratios via AIC (logistic regression).	36
6.4	GLM results and overall model fit.	37
6.5	Accuracy of the SVM model.	39
6.6	Accuracy of the GLM results.	39

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated sources. Where I have consulted the published work of others, in any form (e.g. ideas, equations, figures, text, tables), this is always explicitly attributed.

Berlin, December 24th, 2010

Ceren Önder

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 24.12.2010

Ceren Önder